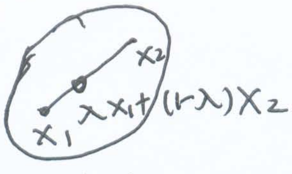
强对偶定理证明

定义1(凸集)：某点集 D 是凸集，是指对于任意两点 $x_1 \in D$，$x_2 \in D$，有：

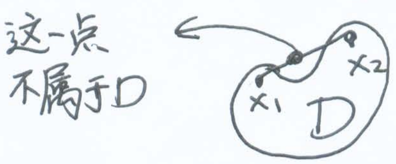
$$x = \lambda x_1 + (1-\lambda) x_2 \in D$$

以下是凸集例子



两点连一条直线，上面所有点都属于此集合。

但下面这个点集不是凸集



定理1(分离超平面定理)：

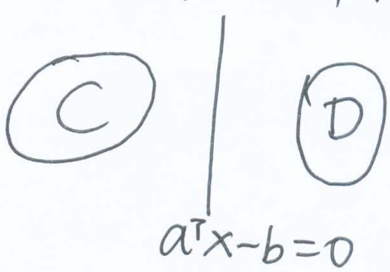
假设两个不相交的凸集 C 和 D，即 $C \cap D = \phi$，则存在向量 $a \neq 0$ 和常数 b，使得对于所有 $x \in C$，有

$$a^T x \leq b$$

且所有 $x \in D$，有

$$a^T x \geq b$$

几何直观解释：两个不相交凸集可以被一个超平面分开。



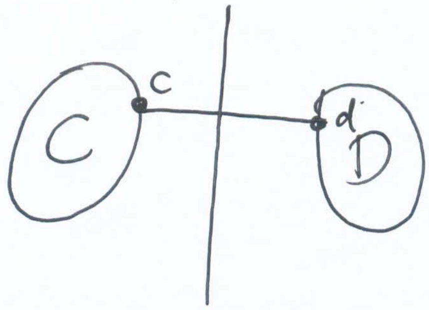
证明：定义点集 C、D 的距离为

$$\text{dist}^2(C, D) = \inf\{\|u-v\|^2 \mid u \in C, v \in D\} \quad \text{(定义2)}$$

并且存在 $c\in C$ 和 $d\in D$ 能达到此最小距离，即

$$\|c-d\|^2 = \mathrm{dist}(C,D)$$

（注：这个假设比较强烈，如果不存在点 c 和 d，我们可以做点集 C 和 D 的闭包，在闭包上获得 c 和 d，同样可以证明。）



$a^Tx-b=0$

定义 $a=c-d$，$b=\dfrac{\|c\|^2-\|d\|^2}{2}$，下面证明

① 对于任意 $u\in C$，有 $a^Tu-b\geq 0$

② 对于任意 $v\in D$，有 $a^Tv-b\leq 0$

只证①，②的证明类似。

反证法：假设存在一个 $u\in C$，使

$a^Tu-b<0$，即

$$(c-d)^Tu-\frac{\|c\|^2-\|d\|^2}{2}<0，即$$

$$(c-d)^T\left(u-\frac{1}{2}(c+d)\right)<0，即$$

$$(c-d)^T\left((u-c)+\frac{1}{2}(c-d)\right)<0，即$$

$$(c-d)^T(u-c)+\frac{1}{2}\|c-d\|^2<0$$

因此有：$(c-d)^T(u-c)<0$

假设另有一点 p 在 u 与 c 的连线上，即

$$p = \lambda u + (1-\lambda)c \text{，其中 } \lambda \in [0,1]$$

根据 C 是凸集，有 $p \in C$，则有：

$$
\begin{aligned}
\|p-d\|^2 &= \|\lambda u + (1-\lambda)c - d\|^2 \\
&= \|(c-d) + \lambda(u-c)\|^2 \\
&= \|c-d\|^2 + 2\lambda(c-d)^T(u-c) + \lambda^2\|u-c\|^2 \\
&= \|c-d\|^2 + \lambda\Big[\underbrace{2(c-d)^T(u-c)}_{\text{此项小于0}} + \underbrace{\lambda\|u-c\|^2}_{\text{此项}\geq 0}\Big]
\end{aligned}
$$

分析一下可知，当 λ 取一个很小的正数时，即

$$\lambda < -\frac{2(c-d)^T(u-c)}{\|u-c\|^2}$$

一定有：$\|p-d\|^2 < \|c-d\|^2$ 且 $p \in C$，这与定义2矛盾。得证。

定理2：若 c 是一个非零向量，即 $\|c\|^2 > 0$，则对任意 $\varepsilon > 0$，存在一个向量 x 满足 ① $\|x\|^2 \leq \varepsilon$；② $c^T x > 0$

证明：取 $x = \dfrac{\varepsilon}{\|c\|^2} c$，则 $\|x\|^2 = \varepsilon$，且

$$c^T x = \varepsilon > 0$$

同理也存在一个向量 x，使 ① $\|x\|^2 \leq \varepsilon$，且 ② $c^T x < 0$

原问题(Prime Problem)

最小化：$f(w)$

限制条件：① $g_i(w) \le 0 \quad (i=1 \sim K)$

② $h_i(w)=0 \quad (i=1 \sim M)$

对偶问题(Dual Problem)

定义：$L(w,\alpha,\beta)$

$$= f(w)+\sum_{i=1}^{K}\alpha_i g_i(w)+\sum_{i=1}^{M}\beta_i h_i(w)$$

最大化：$\theta(\alpha,\beta)=\inf_{(所有w)}\{L(w,\alpha,\beta)\}$

限制条件：$\alpha_i \ge 0 \quad (i=1 \sim K)$

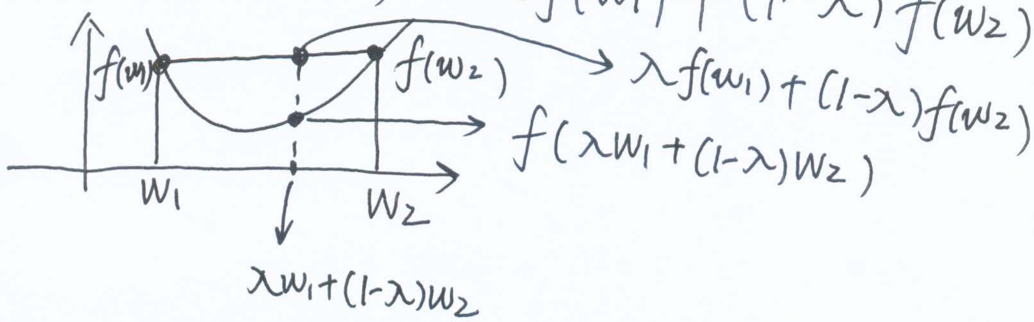
定理3：若 w^* 是原问题的解，(α^*,β^*) 是对偶问题的解，则有：

$$\theta(\alpha^*,\beta^*) \le f(w^*)$$

证明：
$$\theta(\alpha^*,\beta^*)=\inf_{(所有w)}\{L(w,\alpha^*,\beta^*)\}$$
$$\le L(w^*,\alpha^*,\beta^*)$$
$$= f(w^*)+\sum_{i=1}^{K}\underbrace{\alpha_i^*}_{\ge 0}\underbrace{g_i(w^*)}_{\le 0}+\sum_{i=1}^{M}\beta_i^*\underbrace{h_i(w^*)}_{=0}$$
$$\le f(w^*)$$

定义3（凸函数）：$f(w)$是凸函数，是指对 $\forall w_1, w_2$，$\forall \lambda \in [0,1]$，有：

$$f(\lambda w_1 + (1-\lambda) w_2) \le \lambda f(w_1) + (1-\lambda) f(w_2)$$



定理4（强对偶定理）：对于 $f(w)$，$g_i(w)$，$h_i(w)$，

条件1：$f(w)$是凸函数。

条件2：$g_i(w)$是凸函数 $(i=1\sim K)$

条件3：$h_i(w)$是线性函数 $(i=1\sim M)$，即

$$h(w)=\begin{bmatrix} h_1(w) \\ h_2(w) \\ \vdots \\ h_M(w) \end{bmatrix} = \underset{M\times N}{\underline{C}}\ \underset{N\times 1}{\underline{\underline{W}}} + \underset{M\times 1}{\underline{\underline{d}}}$$

条件4（Slater条件）：存在一个w，使

$$g_i(w) < 0 \quad (i=1\sim K)$$

$$h_i(w) = 0 \quad (i=1\sim M)$$

条件5：w的取值范围D是开集，即若 $w \in D$，则存在邻域 $N(w,\varepsilon) \in D$

条件6：w的取值范围D是凸集。

若上述条件满足，则有：

$$f(w^*) = \theta(\alpha^*, \beta^*)$$

证明：构造点集

$$A=\{(u,v,t)\mid \exists w\in D, 使 g_i(w)\le u_i\ (i=1\sim K),\ h_i(w)=v_i\ (i=1\sim M),\ f(w)\le t\}$$

定义：若 $w\in D$，则定义

$$g(w)=\begin{bmatrix} g_1(w)\\ g_2(w)\\ \vdots \\ g_K(w)\end{bmatrix} \quad h(w)=\begin{bmatrix} h_1(w)\\ h_2(w)\\ \vdots \\ h_M(w)\end{bmatrix}$$

注意：

① 若 $w\in D$，则

$(g(w), h(w), f(w))\in A$ （请证明）

② 若 $w\in D$，则

$(+\infty, h(w), +\infty)\in A$ （请证明）

引理1：若D是凸集，$g_i(w)$是凸函数$(i=1\sim k)$
$h_i(w)=cw+d$，$f(w)$是凸函数，则A是凸集。

证明：设$(u_1, v_1, t_1)\in A$且$(u_2, v_2, t_2)\in A$，我们要证$\forall \lambda\in[0,1]$，有

$$(\lambda u_1+(1-\lambda)u_2,\ \lambda v_1+(1-\lambda)v_2,\ \lambda t_1+(1-\lambda)t_2)\in A$$

证明：

$(u_1, v_1, t_1)\in A \Rightarrow \exists w_1\in D$，使
$$g_i(w_1)\le u_{1i}\quad (i=1\sim k)$$
$$h_i(w_1)=v_{1i}\quad (i=1\sim m)$$
$$f(w_1)\le t_1$$

$(u_2, v_2, t_2)\in A \Rightarrow \exists w_2\in D$，使
$$g_i(w_2)\le u_{2i}\quad (i=1\sim k)$$
$$h_i(w_2)=v_{2i}\quad (i=1\sim m)$$
$$f(w_2)\le t_2$$

设$w'=\lambda w_1+(1-\lambda)w_2$，由于$D$是凸集，所以$w'\in D$。
由于$g_i(w)$是凸函数，因此

$$g_i(w')\le \lambda g_i(w_1)+(1-\lambda)g_i(w_2)$$
$$\le \lambda u_{1i}+(1-\lambda)u_{2i}$$

同理有：

$$f(w')\le \lambda t_1+(1-\lambda)t_2$$

而 $h_i(w') = cw' + d$

$$= \lambda(cw_1 + d) + (1-\lambda)(cw_2 + d)$$

$$= \lambda h_i(w_1) + (1-\lambda) h_i(w_2)$$

$$= \lambda v_{1i} + (1-\lambda) v_{2i}$$

命题得证。

根据上述定义，我们有原问题的解

$$f(w^*) = \inf\{t \mid (0,0,t) \in A\}$$

定义另一个点集 $B = \{(0,0,s) \mid s < f(w^*)\}$

可证明 B 也是凸集（略，可作练习）

也可证明：$A \cap B = \phi$（可作练习）

根据定理1（分离超平面定理），存在 (a, β, μ) 使得：

若 $(u,v,t) \in A \Rightarrow a^T u + \beta^T v + \mu t \geq b$（公式1）

若 $(u,v,t) \in B \Rightarrow a^T u + \beta^T v + \mu t < b$

此时 $u=0$ 且 $v=0$，由于有：

$$\mu t < b \quad \text{（公式2）}$$

引理2：若对 $\forall (u,v,t) \in A$，有

$$a^T u + \beta^T v + \mu t \geq b \quad \text{（公式1）}$$

则有：

$$a = \begin{bmatrix} a_1 \geq 0 \\ a_2 \geq 0 \\ \vdots \\ a_k \geq 0 \end{bmatrix} \text{ 且 } \mu \geq 0$$

证明:假设某个 $\alpha_i<0$，则可以取相应 $u_i=+\infty$，此时 (u,v,t) 仍然 $\in A$，则有此时：

$$\alpha^T u+\beta^T v+\mu t=-\infty$$，与公式1矛盾。

同理可证 $\mu\geq 0$

根据A的定义和公式1可得，对 $\forall w\in D$，有:

$$\sum_{i=1}^{K}\alpha_i g_i(w)+\sum_{i=1}^{M}\beta_i h_i(w)+\mu f(w)\geq b$$

因为 $(g(w),h(w),f(w))\in A$

根据B的定义及公式2可得:

$$\mu f(w^*)\leq b$$ (请思考,为何公式2是小于,而这里是小于等于。)

因此有 $\forall w\in D$,

$$\sum_{i=1}^{K}\alpha_i g_i(w)+\sum_{i=1}^{M}\beta_i h_i(w)+\mu f(w)\geq b\geq \mu f(w^*)$$

且 $\alpha_i\geq 0$ $(i=1\sim K)$，$\mu\geq 0$

下面分两种情况讨论

情况1: $\mu\neq 0$，此时有:

$$f(w^*)\leq \sum_{i=1}^{K}\left(\frac{\alpha_i}{\mu}\right)g_i(w)+\sum_{i=1}^{M}\left(\frac{\beta_i}{\mu}\right)h_i(w)+f(w)$$

$$=L\left(w,\frac{\alpha}{\mu},\frac{\beta}{\mu}\right),$$

由于 w 是任意的，因此有

$$f(w^*)\leq \inf_{(所有w)}\left\{L\left(w,\frac{\alpha}{\mu},\frac{\beta}{\mu}\right)\right\}$$，即:

$$f(w^*) \leq \theta\left(\frac{\alpha}{u}, \frac{\beta}{u}\right)$$

由于 $\alpha > 0$，$u > 0$，所以 $\frac{\alpha}{u} > 0$，满足对偶问题限制条件，因此有：

$$f(w^*) \leq \theta(\alpha^*, \beta^*)$$

再根据定理3，有 $\theta(\alpha^*, \beta^*) \leq f(w^*)$ 所以

$$f(w^*) = \theta(\alpha^*, \beta^*)$$，得证。

情况2：$u = 0$，此时有 $\forall w \in D$，有

$$\sum_{i=1}^{K} \alpha_i g_i(w) + \sum_{i=1}^{M} \beta_i h_i(w) \geq 0 \quad (公式3)$$

根据条件4（slater条件），$\exists w$，使

$$g_i(w) < 0 \ (i=1 \sim K) \text{ 且 } h_i(w) = 0 \ (i=1 \sim M)$$

可以推出，$\alpha_i = 0 \ (i=1 \sim K)$，因此公式3变为：

$$\sum_{i=1}^{M} \beta_i h_i(w) \geq 0$$，或

$$\beta^T h(w) \geq 0$$

根据条件3，$h(w) = Cw + d$，代入得：

$$\beta^T (Cw + d) \geq 0$$，即

$$(\beta^T C) w + \beta^T d \geq 0$$

设 $\beta^T C = A$，$\beta^T d = b$，则有

$$Aw + b \geq 0 \quad (公式4)$$

注意公式4对 $\forall w \in D$ 都成立。

根据条件4（Slater条件），∃w，使$Cw+d=0$
即也使 $Aw+b=0$。

下面证明，存在一个 $w'=w+\Delta w$，其中 Δw在w的一个小邻域 $N(0,\varepsilon)$中，使 $Aw'+b<0$。

证明：根据定理1，有 $\beta\neq 0$，否则(a,β,u)都为0，与分离超平面定理矛盾。则有：

$$A=\beta^T C\neq 0$$

根据定理2，存在一个 Δw 满足$\|\Delta w\|^2<\varepsilon$且

$$A\Delta w<0$$

因此：

$$w'=w+\Delta w\in N(w,\varepsilon)。$$

根据条件5， $w'\in D$，同时

$$\begin{aligned}Aw'+b&=A(w+\Delta w)+b\\&=(Aw+b)+A\Delta w\\&=A\Delta w<0\end{aligned}$$

这与公式4矛盾，因此情况2不成立。

命题得证。